My favourite Non-Fiction book of all time is probably Stephen Pinker’s <i>The Better Angels of our Nature</i>. I read that in 2011, and its laboriously detailed demonstration that violence and general human bloody-mindedness has undergone long term decline gave me a fundamentally optimistic outlook on the world that I had never really had before.

Then this month I read <em>Superintelligence</em> and I suspect that that’s all over. This may be the most convincingly depressing Non-Fiction book I can remember reading, and it is difficult to remain optimistic about mankind’s future prospects after reading it.

The superintellignece that the title refers to is a hypothetical future intelligent entity, created by humans, which is by any measure more intelligent than humans. Rostrum spends the first few chapters outlining what is meant by this, and making a convincing case that the development of such an intelligence is highly likely at some point in the coming century.

The second half of the book explains that, unless we are simultaneously very careful, very lucky, and show hitherto undisplayed levels of cooperation, once this genie is let loose we are all doomed.

Most of these dangers, it is argued, come from the sheer omnipotence that such an entity would possess – the running analogy made throughout that the book is that its comparative mental capabilities relative to us would be like ours in comparison with gorillas – conferring similar levels of control over our fate.

I will not list the many traps, dangers and scope for error that are described in the book, but two running themes are:

<ol>

<li>The entitiy need not be malevolent toward humans, it could easily bring about our extinction because we told it do something, and failed to list all the things it should not do in order to achieve it, such as turning the entire planet into computers to work out an answer. It could very easily end our civilisation by accident because we failed to think properly about how to give it goals.</li>

<li>The scope of human suffering possible if we fail to properly deploy superintelligence is extremely large. Such an entity would be capable of creating simulations of human minds so realistic that they would be capable of suffering, which it might well do if we fail to explain why that would be undesirable.</li>

</ul>

A problem with such an account is that after a while everything can sound very fantastical. Despite the great lengths gone to in the beginning to stress the many paths we might take to creating superintelligence, and the rapid speed we are making, you soon stop thinking of this as a possibility because it is so alien to anything we as a species have encountered before. Yet there is no good reason to suppose it will not happen, and many to suppose it will. There is a danger of dismissing the risks in the same way that many people dismiss those from climate change – accepting it as real on an academic level, but not really internalising the danger on a gut level.

Some words of advice are offered throughout, particularly in the field of ‘motivation selection’ – choosing how to give the superintellignece a set of values that might avert doom. Much if it sounds hopelessly optimistic, particularly in light of the sheer magnitude of the advantages of that being the first to develop such an entity could offer if done right. There will be little incentive to slow down and think about safety.

As is often the case with good audiobooks, the narrator is superb. Napoleon Ryan’s delivery is one of the best I’ve heard.

It’s probably impossible to give an overall sentiment about such a book as this so soon after reading it, so I’m not sure what my closing thoughts are to be. It may be that once the dust has cleared, the dire warnings of this book seem much less pressing or threatening as they do in its immediate aftermath. But given the rapid pace of AI development at the current time, and the increasing frequency with which it makes the news, it may be a topic that is impossible to forget about, for better or for worse.

Read – December 2015/January 2016